

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
ELEKTROTEHNIČKI FAKULTET

Sveučilišni studij

ALGORITAM DIFERENCIJALNE EVOLUCIJE ZA
AUTOMATSKO GRUPIRANJE PODATAKA

Diplomski rad

Karlo Kotrba

Osijek, 2015.

SADRŽAJ

1.	UVOD	1
2.	GRUPIRANJE PODATAKA	3
2.1.	Čvrsto grupiranje podataka.....	3
2.2.	Algoritmi za grupiranje podataka	4
2.3.	Određivanje broja grupa	6
2.4.	Primjena grupiranja podataka.....	10
3.	DIFERENCIJALNA EVOLUCIJA	12
3.1.	Način rada diferencijalne evolucije	12
3.2.	Primjena za grupiranje podataka	16
3.3.	Primjena za automatsko grupiranje podataka.....	18
4.	OSTVARENO PROGRAMSKO RJEŠENJE.....	20
4.1.	Izvedba programskog rješenja	20
4.2.	Prikaz i način uporabe programskog rješenja.....	30
5.	EKSPERIMENTALNA ANALIZA	33
5.1.	Postavke eksperimenta	34
5.2.	Rezultati eksperimentalne analize	35
6.	ZAKLJUČAK	43

LITERATURA

SAŽETAK

ABSTRACT

ŽIVOTOPIS

PRILOZI (NA CD-U)

SAŽETAK

Diplomski rad obrađuje uvod u algoritme za grupiranje podataka. Ukratko su obrađeni glavni tipovi algoritama, te načini vrednovanja kvalitete particije. Detaljnije je objašnjena diferencijalna evolucija uz povezivanje sa evolucijskim algoritmima. Prikazani su načini primjene diferencijalne evolucije pri grupiranju podataka u poznat ili nepoznat broj grupa. Također je opisan način pronalaženja odgovarajućeg broja grupa ukoliko on nije unaprijed poznat. Izrađeno je programsko rješenje koje omogućuje usporedbu raznih načina vrednovanja kvalitete grupiranja. Također je provedena eksperimentalna analiza gdje su prikazani i komentirani rezultati usporedbe. Uz ovo analizirana je i efikasnost raznih načina paraleliziranja programskog koda.

Ključne riječi: automatsko grupiranje podataka, diferencijalna evolucija, grupiranje podataka, kriteriji vrednovanja grupiranja, paralelne petlje.

ABSTRACT

The thesis gives an introduction to data clustering algorithms. Main types of data clustering algorithms as well as different clustering validity criteria were described. Differential evolution and its application related to data clustering was described in detail. The applicability of differential evolution for automatic clustering as well as clustering when the number of clusters is known in advance was shown. Automatic clustering was described in depth. Software solution that enables a comparison between different clustering validity criteria was developed. The results of the aforementioned comparison were shown and commented. Efficacy analysis of multiple implementations of parallelism was also shown.

Key words: automatic data clustering, differential evolution, data clustering, cluster validity index, parallel loops.